The goal of this project is to develop the fundamental knowledge necessary to engineer anaerobic gut fungi as novel platform organisms for lignocellulosic biofuel production. Anaerobic fungi are the primary colonizers of biomass within the digestive tract of large herbivores, where they have evolved unique abilities to break down lignin-rich cellulosic biomass through invasive, filamentous growth and the secretion of powerful lignocellulolytic enzymes and enzyme complexes (cellulosomes). Despite these attractive abilities, considerably less genomic and metabolic data exists for gut fungi compared to well-studied anaerobic bacteria and aerobic fungi that hydrolyze cellulose. This presents a significant knowledge gap in understanding gut fungal function, substrate utilization, and metabolic flux, which has prohibited the genetic and functional modification of gut fungi. Recently, however, advances in sequencing technologies have made it possible to explore the dynamic metabolic networks within gut fungi for the first time. Our approach combines next-generation sequencing with physiological characterization to establish the critical knowledge base to understand lignocellulose breakdown by gut fungi. This approach will enable exploration of novel isolates for desirable enzymatic properties, construction of metabolic models to describe biomass degradation, and new methods to engineer gut fungi for bioprocessing. Specific aims to be completed during the project are detailed below.

Aim 1: Conduct a functional screen of gut fungal isolates from herbivores that consume grass forage, and those that consume woody biomass to identify optimal gut fungi for lignocellulose breakdown. Correlate biomass-degrading capacity with microbe morphology and enzymatic diversity, and submit desirable isolates for further genomic characterization by the DOE-JGI.

Aim 2: Construct metabolic models to describe transcriptional mechanisms responsible for enzyme regulation in the gut fungus *Piromyces sp finn*. Quantify global changes in the fungal transcriptome when simple sugars repress lignocellulosic activity (catabolic regulation).

Aim 3: Develop enabling technologies for the transformation and genetic reprogramming of the gut fungus *Piromyces sp finn*. Proof of concept will be achieved through transformation with a gene encoding fluorescent protein (FbFP), with later applications to focus on engineering substrate utilization in fungi by eliminating metabolic bottlenecks.

The following report details major insights, discoveries, and accomplishments under each aim of the project, which resulted in 19+ peer-reviewed publications.

Aim 1: Genomic Sequencing Allows Characterization of Fungal Enzymes and Cellulosomes

At the outset of this project, a fungal enrichment experiment was conducted by the O'Malley Lab to isolate biomass-degrading fungi from large herbivores at the Santa Barbara Zoo that would serve as model systems. These isolates have officially been assigned species names, and are indexed in the Index Fungorum (http://www.indexfungorum.org). In the past 4 years, our group has worked with collaborators at the Broad Institute of MIT and Harvard, as well as the Joint Genome Institute (JGI) to characterize the genomic, transcriptomic, and proteomic machinery of these early-branching fungi. As part of our recent publication in *Science*, we found that the number of biomass-degrading genes in the anaerobic fungi was richer than any other microbe yet described in nature (Figure 1). Compared to "higher" (e.g. more evolved fungi), the more primitive anaerobic fungi are equipped with a rather even distribution of cellulose-degrading enzymes, xylan-degrading enzymes, and accessory enzymes that have been largely lost to

evolution. This illustrates not only their powerful biomass-degrading abilities, but also the potential to identify completely novel sequences in their genomes that could be exploited for biotechnology.

Our group has worked with collaborators at the Joint Genome Institute (JGI) (via approved CSP projects) to sequence genomes of the anaerobic fungi *Neocallimastix californiae*, *Anaeromyces robustus*, *Caecomyces churrovis*, and *Piromyces finnis*. Our recent publications in Science and Nature Microbiology released the first high-quality transcriptomes and genomes of the anaerobic gut fungi (results available on Mycocosm). With the advent of PacBio sequencing, long reads averaging 10,000bp in length were sufficient to remedy the assembly of highly repeat-rich regions of the genomes that could not be resolved with Illumina-based approaches. These results also enabled further epigenetic analysis of the fungal genomes (ongoing collaboration with Igor Grigoriev's group at JGI), that highlighted regions of the genomes that are highly transcribed (manuscript recently published in Nature Genetics).

In previous years of this project, we established that gut fungi secrete a large number of cellulolytic enzymes, which form cellulosomes that may be physically attached to the cell (Fig. 2). However, the enzyme components, modular assembly mechanism, and functional role of fungal cellulosomes during biomass breakdown remained unknown. Although the basic interaction of these complexes is a modular cohesion-dockerin binding similar to that observed in bacteria, the fungal dockerin and scaffoldin domains have no similarity to their bacterial counterpart. Previously obtained genomic and transcriptomic data for three fungal strains isolated by our lab enabled us to identify a "parts list" for fungal cellulosomes, including a set of novel scaffolding proteins that biochemically interact with dockerin-fused enzymes from fungi (Haitjema, 2017 *Nature Microbiology*). Further, our previous work has revealed that (i) only some of the scaffoldin proteins have putative transmembrane helix domains, (ii) many of the dockerin domain proteins (DDPs) do not identify as carbohydrate active enzymes (CAZy) and (iii) fungi still secrete up to 50% of their CAZymes as free enzymes (lacking a dockerin domain) (Haitjema, 2017 *Nature Microbiolgy*). Finally, comparative genomics against anaerobic bacteria and other strains of anaerobic fungi revealed that a number of CAZyme domains identified in the fungal genomes originated from bacteria (Figure 2). This finding indicates that fungi and bacteria that co-exist in the rumen of large herbivores may have exchanged genetic information, allowing for the optimization of biomass degradation in these anaerobic systems. In particular, this finding is surprising, as horizontal gene transfer in this respect is typically observed to occur within life forms that share the same kingdom of life – it is seldom described to occur from prokaryotes to eukaryotes.

Aim 2: Metabolic Modeling of Anaerobic Fungi Provide a Path Forward for CBP

Compared to model organisms, there is a relative dearth of information about the metabolism of anaerobic gut fungi. To address this issue, we used transcriptomic information to identify enzymes from critical metabolic processes for the first time. Transcripts identified in the *de novo* assembled transcriptomes were functionally annotated using a combination of NCBI BLAST, EMBL-EBI InterProScan, and Gene Ontology mapping. Enzymes were identified by their Enzyme Commission (EC) numbers, which were only assigned to transcripts with a high similarity to

InterPro protein domains. Combining this information with metabolic maps from KEGG databases, the core metabolic maps were filled in and the sugars that could be metabolized by the fungus were identified. This analysis revealed that the fungus contains all of the enzymes required for glycolysis through ethanol fermentation, but takes an irreversible route as it is missing phosphoglycerate mutase (EC:5.4.2.11), which indicates that the gut fungi cannot perform gluconeogenesis. Furthermore, all enzymes necessary for the metabolism of fructose, mannose, sucrose, α-galactose, and xylose were present. Enzymes required for the metabolism of arabinose and β-galactose were missing from the transcriptome. Growth experiments have previously shown that the gut fungi cannot grow on xylose or galactose despite the presence of the enzymes required to metabolize them. This may indicate a limitation due to the transport of the sugars into the cell rather than in the ability to metabolize them. Growth was not observed on arabinose, which supports the claim that the necessary enzymes are not present. Natural metabolic products of biomass fermentation by anaerobic gut fungi include ethanol, lactate, formate, acetate, carbon dioxide, and hydrogen. These products can facilitate the link between the fungi and other organisms in their native environment (e.g. methanogens most commonly consume the CO_2 and H_2 produced and reduce them further to methane).

A longer-term goal of our work is to not only identify novel gut fungal enzymes, to employ these organisms as platforms for biomass breakdown and conversion. While genetic tools continue to be developed for these recalcitrant fungi, we have in the meantime developed strategies to co-culture them with model microbes. In this way, anaerobic fungi are used to consolidate pretreatment and hydrolysis steps, while facultative anaerobes are used to funnel sugar-rich hydrolysates to value added products. In the past two years, we developed a biphasic fermentation scheme that combines the lignocellulolytic action of anaerobic fungi isolated from large herbivores with domesticated microbes for bioproduction. When grown in batch culture, anaerobic fungi release excess sugars from both cellulose and crude biomass due to a wealth of highly expressed carbohydrate active enzymes (CAZymes), converting as much as 49% of cellulose to free glucose (Figure 3). This sugar-rich hydrolysate readily supports growth of *S. cerevisiae*, which can be engineered to produce a range of value-added chemicals (Figure 3). Further, reconstruction of metabolic pathways from transcriptomic data reveals that anaerobic fungi do not catabolize all sugars that their enzymes hydrolyze from biomass, leaving other carbohydrates such as galactose, arabinose, and mannose available as nutritional links to other microbes in their consortium (Figure 3); this work was recently published in Biotechnology and Bioengineering. Although basal expression of CAZymes in anaerobic fungi is high, it is drastically amplified by cellobiose breakout products encountered during biomass hydrolysis. Overall, these results suggest that anaerobic fungi provide a nutritional benefit to the rumen microbiome, which can be harnessed to design synthetic microbial communities that compartmentalize biomass degradation and bioproduct formation.

Aim 3: Genetic Engineering of Anaerobic Fungi

Genetic modification of gut fungi will allow us to overexpress, purify, and test the function of enzymes directly in their native host. Additionally, this approach guarantees proper protein biogenesis (and post-translational modification) machinery that may be required for fungal cellulosome expression and activity. In previous years of the project, we have taken the first

major steps towards a method to stably genetically transform anaerobic gut fungi, *which has never been successfully reported in the literature.* Beyond the obvious lack of knowledge of their genomes, the very thick cell wall of anaerobic fungi (greater than 5um) often prevents delivery of any amount of DNA. So far, we have shown that gut fungal zoospores (cells in the early stage of their life cycle) are amenable to transformation by electroporation (DNA uptake during electroshock). This optimized procedure was the result of a design of experiments approach, where electrotransformation, biolistic transformation, and chemical transformation methods were evaluated as applied to fungal zoospores and mature zoosporangium. Efforts over the past year have largely focused on leveraging genomic information to (1) identify putative regulatory regions (e.g. promoters/terminators) within gut fungal genomes, and (2) narrow in on highly accessible regions of the genome to facilitate high-efficiency homologous recombination of foreign DNA. In addition, we are building the components of a CRISPR/Cas9 genomic editing strategy that can be applicable to the anaerobic fungi. To date, we have a catalog of putative promoter elements, and are now focusing on transcriptionally-active long-transcribed repeat regions of the fungal genomes as target destinations for genentic engineering.

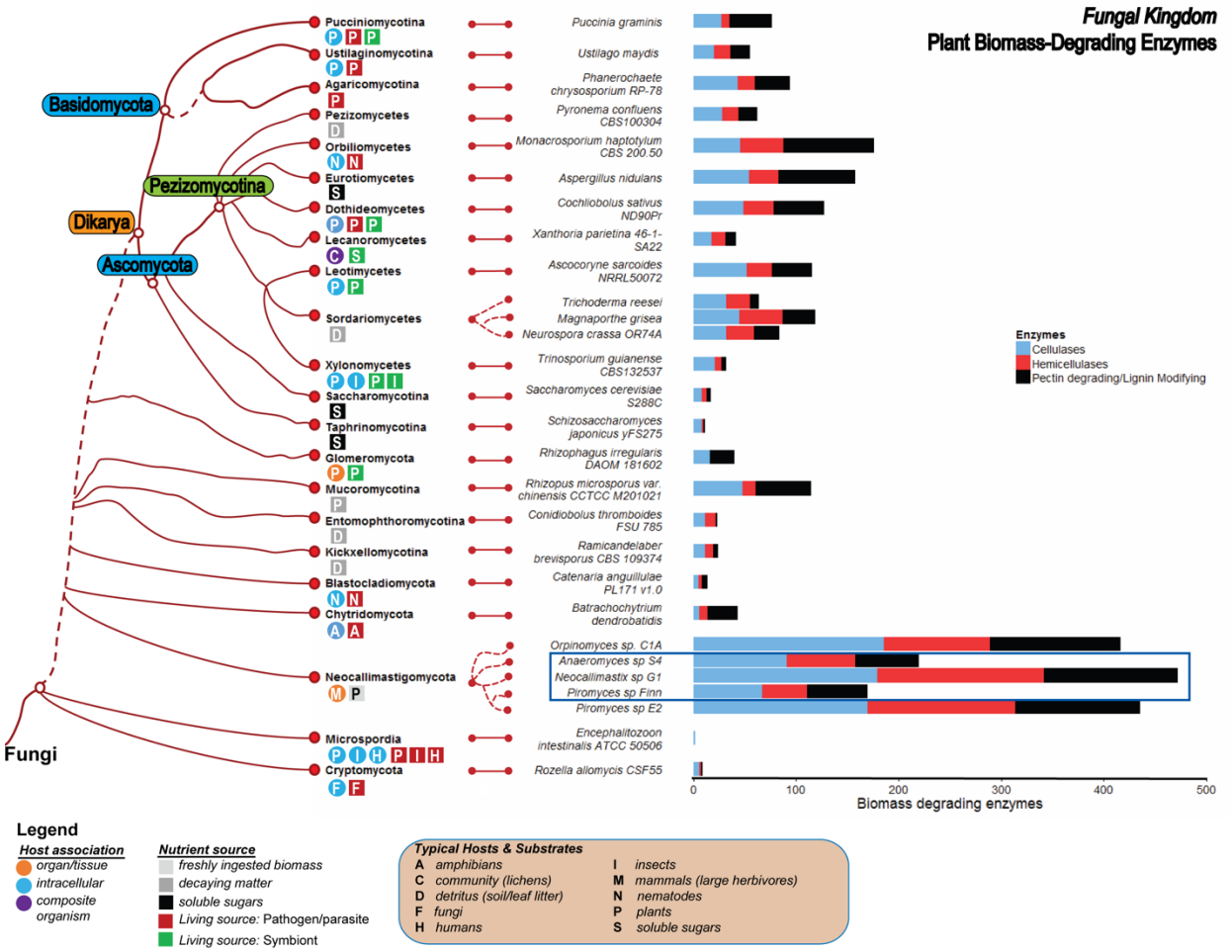

Fig. 1| Biomass degrading machinery in the fungal kingdom. Biomass degrading genes (Table S1) within the genomes of representative fungal species. Boxed species were isolated and their transcriptomes sequenced in this paper (Database S1-S3). Gene numbers for these isolates are estimated from the transcriptome. Fungal Tree of Life adapted from that at Mycocosm (*14*). Common host associations and substrate preferences are indicated below each fungal division.

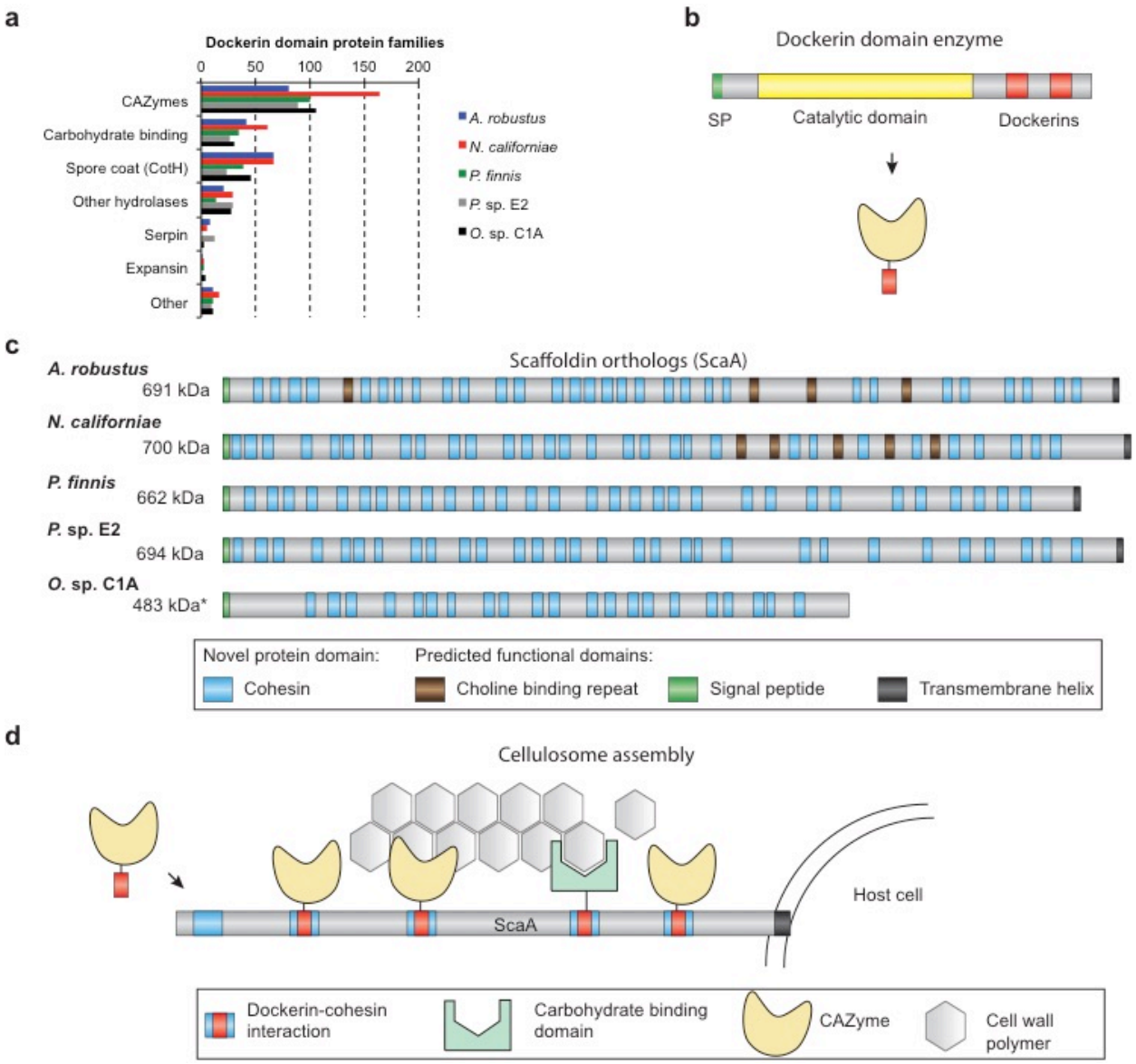

Figure 2. Overview of gut fungal cellulosome components. a, Protein functional domains associated with non-catalytic dockerin domains. **b,** Schematic of a typical dockerin domain glycoside hydrolase (GH). **c,** Schematic of large non-catalytic scaffoldin molecules in the cellulosomes of gut fungi. The predicted functional domains were determined by InterProScan 5. The extracellular domains (shaded grey) are decorated with interspersed repeating cohesin motifs (shaded blue). The predicted N-terminal signal sequence and C-terminal membrane helix are shaded in green and black, respectively. **d,** A cartoon model of gut fungal cellulosome assembly.
*The sequence of the *O*. sp. C1A *ScaA* gene model is incomplete due to fragmented genome assembly.

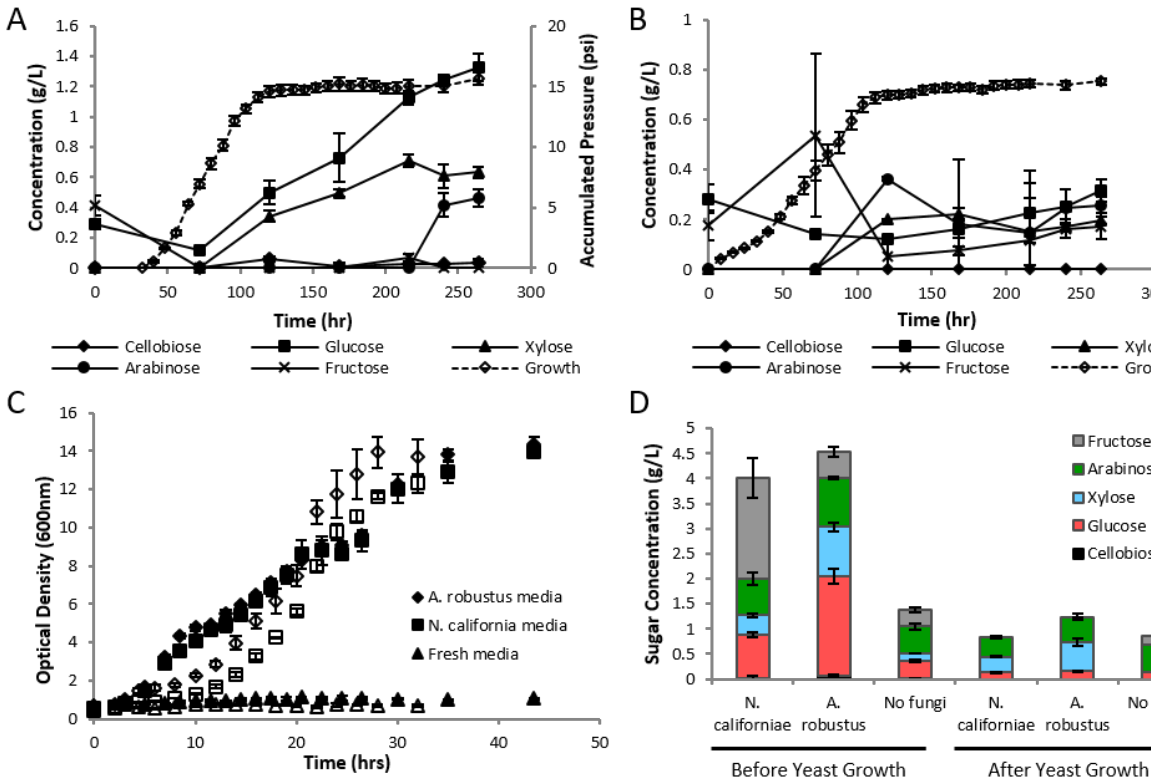

Figure 3. Excess sugars are released from cellulosic and lignocellulosic substrates by anaerobic fungi. A) Growth of A. robustus on 0.5 g of reed canary grass in 10 mL culture, and sugar cocentrations released from biomass. Growth (pressure) data is shown in empty symbols and sugar data in solid symbols. B) Growth of N. californiae on 0.5 g of reed canary grass in 10 mL culture, and sugar concentrations released from biomass. Growth (pressure) data is shown in empty symbols and sugar data in solid symbols. C) Growth of S. cerevisiae on fungal spent media. Spent media containing crystalline cellulose broken down by the fungi into glucose (filled symbols) or reed canary grass broken down into glucose and other sugars (empty symbols). D) End point sugar concentrations produced after fungal growth on reed canary grass, and sugar concentration after yeast proliferation in spent fungal hydrolysate media.